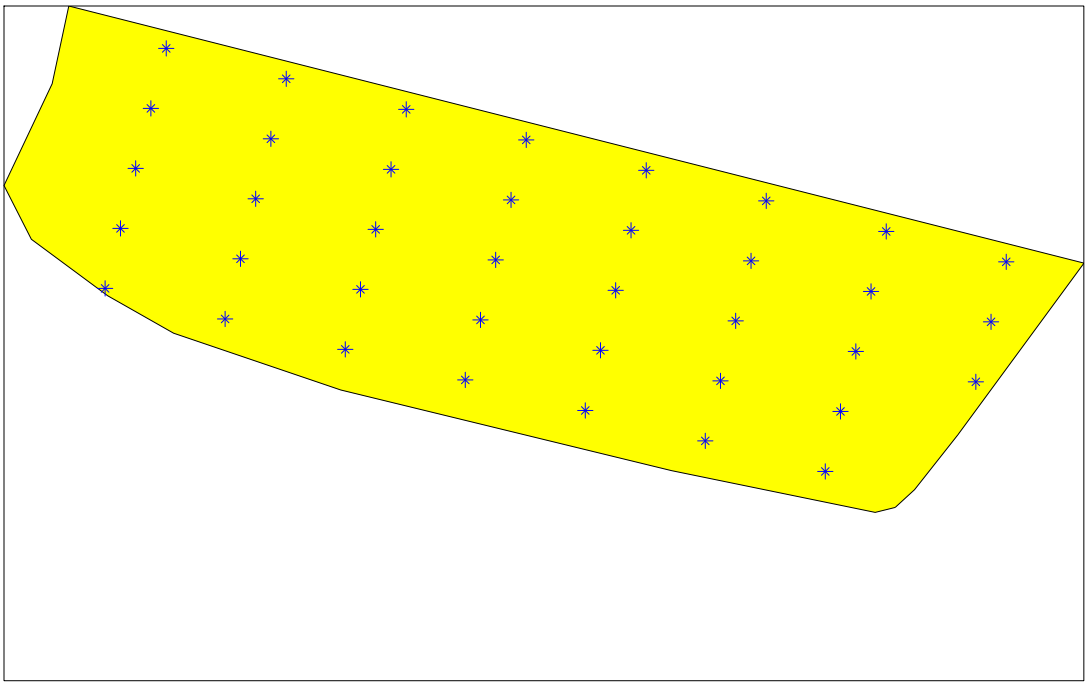
Systematic sampling locations for detecting an area of elevated values (hot spot)

This report summarizes the sampling design used, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (e.g., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design developed. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Detect the presence of a hot spot that has a specified size and shape
Type of Sampling Design	Hot spot
Sample Placement (Location) in the Field	Systematic (Hot Spot) with a random start location
Formula for calculating minimum size of hot spot	Algorithm developed by Singer and Wickman (1969)
Calculated total number of samples	38
Type of samples	Point Samples
Number of samples on map [a]	38
Number of selected sample areas [b]	1
Specified sampling area [c]	2035162.35 ft^2
Grid pattern	Rectangular
Size of grid / Area of grid [d]	50 x 100 meters / 53819.6 m^2
Total cost of sampling [e]	$20,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Size of grid / Area of grid gives the linear and square dimensions of the grid spacing used to systematically place samples.
[e] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	**Y Coord**	**Label**	**Value**	**Type**	**Historical**
2845684.7876	2341166.9729			Hotspot	
2846002.7587	2341086.1428			Hotspot	
2846320.7298	2341005.3128			Hotspot	
2846638.7009	2340924.4828			Hotspot	
2846956.6720	2340843.6527			Hotspot	
2847274.6430	2340762.8227			Hotspot	
2847592.6141	2340681.9927			Hotspot	
2845725.2026	2341325.9584			Hotspot	
2846043.1737	2341245.1284			Hotspot	
2846361.1448	2341164.2983			Hotspot	
2846679.1159	2341083.4683			Hotspot	
2846997.0870	2341002.6383			Hotspot	
2847315.0581	2340921.8082			Hotspot	
2847633.0291	2340840.9782			Hotspot	
2845765.6176	2341484.9440			Hotspot	
2846083.5887	2341404.1139			Hotspot	
2846401.5598	2341323.2839			Hotspot	
2846719.5309	2341242.4539			Hotspot	
2847037.5020	2341161.6238			Hotspot	
2847355.4731	2341080.7938			Hotspot	
2847673.4442	2340999.9638			Hotspot	
2847991.4152	2340919.1337			Hotspot	

2845806.0327	2341643.9295			Hotspot	
2846124.0037	2341563.0995			Hotspot	
2846441.9748	2341482.2694			Hotspot	
2846759.9459	2341401.4394			Hotspot	
2847077.9170	2341320.6094			Hotspot	
2847395.8881	2341239.7793			Hotspot	
2847713.8592	2341158.9493			Hotspot	
2848031.8303	2341078.1193			Hotspot	
2845846.4477	2341802.9150			Hotspot	
2846164.4188	2341722.0850			Hotspot	
2846482.3898	2341641.2550			Hotspot	
2846800.3609	2341560.4249			Hotspot	
2847118.3320	2341479.5949			Hotspot	
2847436.3031	2341398.7649			Hotspot	
2847754.2742	2341317.9348			Hotspot	
2848072.2453	2341237.1048			Hotspot	

Primary Sampling Objective
The primary purpose of sampling at this site is to detect "hot spots" (local areas of elevated concentration) of a given size and shape with a specified probability, 1-β.

Selected Sampling Approach
This sampling approach requires systematic grid sampling with a random start. If a systematic grid is not used, the probability of detecting a hot spot of a given size and shape will be different than desired or calculated.

Number of Total Samples: Calculation Equation and Inputs
The algorithm used to calculate the probability of a hit (which makes possible the calculation of the hot spot size or the number of samples) was developed by Singer and Wickman (1969) and Singer (1972) with refinements by Davidson (1995). Gilbert (1987) also discussed hotspot sampling designs. Inputs to the algorithm include the size, shape, and orientation of a hot spot of interest, an acceptable probability of finding a hot spot, the desired type of sampling grid, and the sampling budget. For this design, the smallest hot spot that could be detected was calculated based on the given grid size and other parameters.

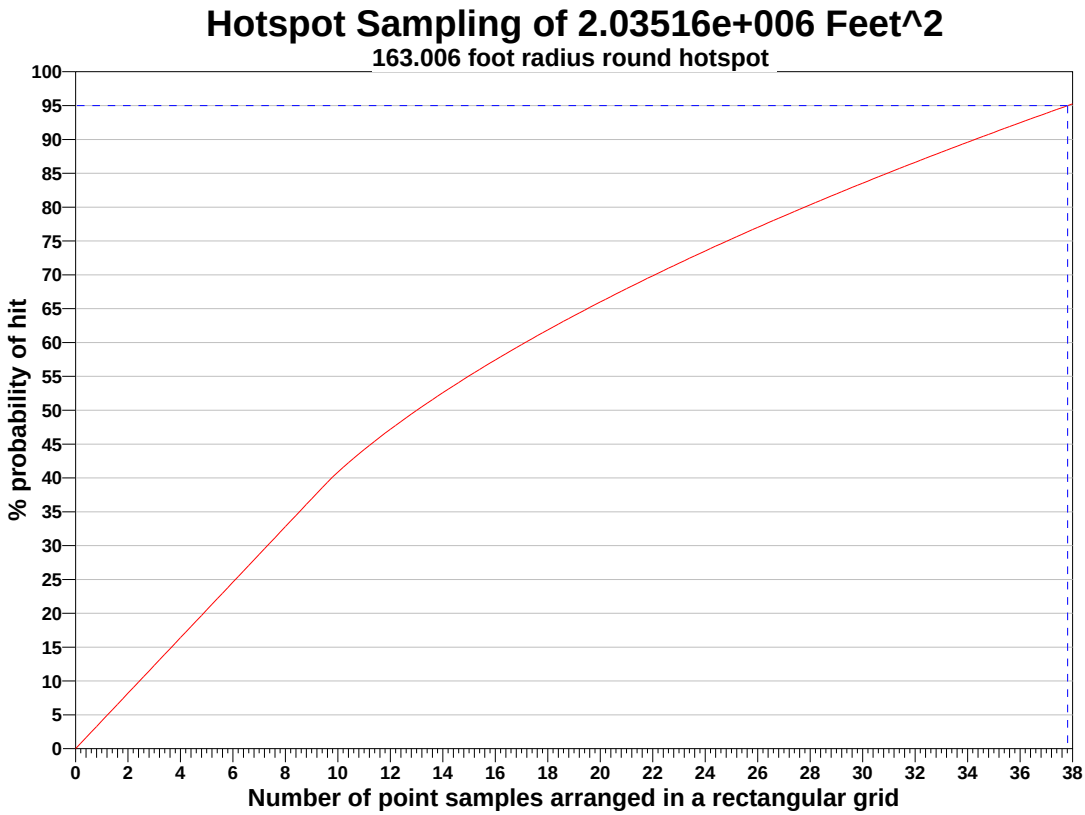
The inputs to the algorithm that result in the smallest hot spot that could be detected are:

Parameter	**Description**	**Value**
Inputs		
1-β	Probability of detection	95%
Grid Type	Grid pattern (Square, Triangular or Rectangular)	Rectangular
Rectangle Ratio	Rectangle width to height ratio	2
Grid Size	Spacing between samples	50 x 100 meters
Grid Area	Area represented by one grid	53819.6 m^2
Sample Type	Point samples or square cells	Points
Hot Spot Shape	Hot spot height to width ratio	1
Angle	Angle of orientation between hot spot and grid	Random
Sampling Area	Total area to sample	2035162.35 ft^2
Outputs		

Hot Spot Size	Length of hot spot semi-major axis	163.006 feet
Hot Spot Area [a]	Area of hot spot (Length^2 * Shape * π)	83474.8 ft^2

[a] Length of semi-major axis is used by Singer-Wickman algorithm. Hot spot area is provided for informational purposes.

The following graph shows the relationship between the number of samples and the probability of finding the hot spot. The dashed blue line shows the actual number of samples for this design (which may differ from the optimum number of samples because of edge effects).



Assumptions that Underlie the VSP Locating a Hot Spot Design Method

1. In the decision area there is at least one hotspot of the designated size, which is circular or elliptical in shape.
2. The level of contamination that defines a hotspot is well defined.
3. The location of the hotspot is unknown, and if a hotspot is present, all locations within the sampling area are equally likely to contain the hotspot.
4. With a randomly determined starting location, samples are taken on a square, rectangular or triangular (equilateral) grid pattern that covers the decision area.
5. Each sample is collected, handled, measured or inspected using approved methods that yield sufficiently precise measurements.
6. A very small proportion of the surface of the decision area will be sampled. The area sampled by a single sample is much smaller than the hotspot of interest.
7. The sample methodology and sample analysis process is the same for all sample locations.
8. There are no classification errors. If a hotspot is sampled, then contamination is detected (i.e., no false negatives). If an uncontaminated area is sampled, it is not mistakenly identified as a hotspot (i.e., no false positives).

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the probability of hit (%), hot spot shape (height to width ratio) and hot spot size (length of semi-major axis). The following table shows the results of this analysis.

Number of Samples

		Size=81.5029	Size=163.006	Size=244.509
1-β=90	Shp=0.8	38	38	38
	Shp=0.9	38	38	38
	Shp=1	38	38	38
1-β=95	Shp=0.8	38	38	38
	Shp=0.9	38	38	38
	Shp=1	38	38	38
1-β=100	Shp=0.8	38	38	38
	Shp=0.9	38	38	38
	Shp=1	38	38	38

1-β = Probability of Hit (%)
Shp = Hot Spot Shape (Height to Width Ratio)
Size = Hot Spot Size (Length of Semi-major Axis)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $20,000.00, which averages out to a per sample cost of $526.32. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**38 Samples**
Field collection costs		$100.00	$3,800.00
Analytical costs	$400.00	$400.00	$15,200.00
Sum of Field & Analytical costs		**$500.00**	**$19,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$20,000.00**

Recommended Data Analysis Activities
Post data collection activities generally follow those outlined in EPA's Guidance for Data Quality Assessment (EPA, 2006). The data analysts will become familiar with the context of the problem and goals for data collection and assessment. The data will be verified and validated before being subjected to statistical or other analyses. Graphical and analytical tools will be used to verify to the extent possible the assumptions of any statistical analyses that are performed as well as to achieve a general understanding of the data. The data will be assessed to determine whether they are adequate in both quality and quantity to support the primary objective of sampling.

A map of the actual sample locations will be generated so that the sampling plan and the field implementation may be compared. Deviations from planned sample locations due to topographic, vegetative, or other features will be noted. Their impacts will be qualitatively assessed. If a hot spot is discovered, additional sampling may be performed to determine its size and shape, in which case, the initial assumptions of the sampling design may then be assessed and/or reconsidered.

References
EPA 2006. *Data Quality Assessment: Statistical Methods for Practitioners EPA QA/G-9S*, EPA/240/B-06/003, U.S. Environmental Protection Agency, Office of Environmental Information, Washington DC.

Davidson, J.R. 1995. *ELIPGRID-PC: Upgraded Version.* ORNL/TM-13103. Oak Ridge National Laboratory, Oak Ridge, TN.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Wiley & Sons, Inc., New York, NY.

Singer, D.A. and J.E. Wickman. 1969. *Probability Tables for Locating Elliptical Targets with Square, Rectangular, and Hexagonal Point Nets*. Pennsylvania State University, University Park, Pennsylvania. Special Publication 1-69.

Singer, D.A. 1972. "ELIPGRID: A Fortran IV program for calculating the probability of success in locating elliptical targets with square, rectangular and hexagonal grids." *Geocom Bulletin/Programs* 4:1-16.